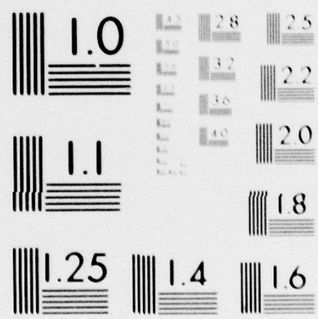
AD-A078 602 BROWN UNIV PROVIDENCE R I LEFSCHETZ CENTER FOR DYNAM--ETC F/G 12/1
A PROJECTED STOCHASTIC APPROXIMATION METHOD FOR ADAPTIVE FILTER--ETC(U)
AUG 79 H J KUSHNER N00014-76-C-0279

UNCLASSIFIED AFOSR-TR-79-1250 NL

1 OF 1

AD
A078602

END
DATE
FILMED
1 -80
DDC


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) unclassified

(2)

LEVEL II

ADA078602

DDC FILE COPY

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER (18) AFOSR-TR- 79-1250	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) A PROJECTED STOCHASTIC APPROXIMATION METHOD FOR ADAPTIVE FILTERS AND IDENTIFIERS		5. TYPE OF REPORT & PERIOD COVERED (9) Interim rept.
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) Harold J. Kushner		8. CONTRACT OR GRANT NUMBER(s) AFOSR-77-3063
9. PERFORMING ORGANIZATION NAME AND ADDRESS Brown University, Division of Applied Mathematics, Providence, RI 02912		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A1
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM, Bolling AFB, Washington, DC 20332		12. REPORT DATE (11) 22 Aug[redacted]79
		13. NUMBER OF PAGES 13
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (12) 15p		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

(16) 2304

(17) A1

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

(15) N00014-76-C-0279-, AFOSR-77-3063

DDC
RECEIVED
DEC 18 1979
B

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Generally, when stochastic approximation is used to identify the coefficients of a linear system or for an adaptive filter or equalizer, the iterate X_n is projected back onto some finite set $G = \{x: |x_i| \le B, \text{ all } i\}$, if it ever leaves it. The convergence of such truncated sequences have been discussed informally. Here it is shown, under very broad conditions on the

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

JOB

401 834

20. Abstract continued.

noises, that $\{X_n\}$ converges w.p.1. to the closest point in G to the optimum value of X_n. Also, under even weaker conditions, the case of constant coefficient sequence is treated, and a weak convergence result obtained. The set G is used for simplicity. It can be seen that the result holds true in more general cases but the box is used since it is the only commonly used constraint set for this problem.

ACCESSION for		
NTIS	White Section	☑
DDC	Buff Section	☐
UNANNOUNCED		☐
JUSTIFICATION		
BY		
DISTRIBUTION/AVAILABILITY CODES		
Dist. AVAIL. and/or SPECIAL		
A		

AFOSR-TR- 79-1250

A PROJECTED STOCHASTIC APPROXIMATION METHOD FOR ADAPTIVE FILTERS AND IDENTIFIERS

Harold J. Kushner[+]

Fellow, IEEE

Abstract

Generally, when stochastic approximation is used to identify the coefficients of a linear system or for an adaptive filter or equalizer, the iterate X_n is projected back onto some finite set $G = \{x: |x_i| \leq B, \text{ all } i\}$, if it ever leaves it. The convergence of such truncated sequences have been discussed informally. Here it is shown, under very broad conditions on the noises, that $\{X_n\}$ converges w.p.1. to the closest point in G to the optimum value of X_n. Also, under even weaker conditions, the case of constant coefficient sequence is treated, and a weak convergence result obtained. The set G is used for simplicity. It can be seen that the result holds true in more general cases but the box is used since it is the only commonly used constraint set for this problem.

[+]Brown University, Divisions of Applied Mathematics and Engineering and Lefschetz Center for Dynamical Systems. Research supported in part by the Air Force Office of Scientific Research under AFOSR AF-77-3063, by the National Science Foundation under NSF Eng. 77-12946, and in part by the Office of Naval Research ONR N0014-76-C-0279-P0002.

August 22, 1979

This report

Introduction

~~Reference [1]~~ deals with a great variety of stochastic approximation procedures, for constrained and unconstrained systems and for convergence w.p.1. and weak convergence, all for systems with correlated inputs. The techniques ~~of [1]~~ are readily usable for many problems that are not explicitly treated there. This will be illustrated here for one particular class of constrained problems which is of great current interest and which arises in identification and in adaptive control theory. In fact, it is just such constrained problems to which more attention should be given, owing to their prevalence. The proofs are contained in various parts ~~of [1]~~ and, here, after the problem is defined, it is shown how to put the bits and pieces together. The The problem and method are typical of a large class of adaptive systems which can be treated by similar methods, and is worth singling out.

The problem will be set up in such a way that it fits both a standard identification problem and a standard problem in adaptive equalizers. Let $\{d_n\}$ denote a scalar valued desired output sequence, perhaps a training or reference signal, or output of the system to be identified. The problem can readily be set up so that all quantities $\{d_n, u_n, X_{ni}, \rho_n\}$ are complex valued, but in the interest of simplicity, we suppose that they are real valued. Let $\{u_n\}$ denote an input sequence set $\psi_n \equiv (u_n, \ldots, u_{n-r+1})'$, and let $\{\rho_n\}$ be a noise sequence, independent of $\{u_n\}$. The observed adaptive system output at time n is defined by

$\sum_{i=0}^{r-1} X_{ni}u_{n-i} = y_n$, and the 'perturbed' observed reference at time n is $d_n + \rho_n$. The idea in [2], [3], [4] and in many other papers is to adjust the system parameter $X_n = (X_{n0},\dots,X_{n,r-1})'$ so that the output $\{y_n\}$ 'best matches' the $\{d_n\}$ in a mean square sense. A common recursive adaptive algorithm for doing this is

$$
\begin{aligned}
X_{n+1} &= X_n - a_n\psi_n\varepsilon_n, \quad \varepsilon_n = (y_n - d_n - \rho_n) \\
&= X_n - a_n\psi_n(\psi_n' X_n - d_n - \rho_n), \quad a_n \to 0, \quad \sum_n a_n = \infty, \quad a_n > 0.
\end{aligned} \tag{1.1}
$$

Algorithm (1.1) has been the focus of an enormous amount of effort. In practice, there is usually given a bound B such that if some $|X_{ni}| > B$, then X_{ni} is immediately reset to the closest value $+B$ or $-B$. This projected version has received little attention. Ljung [2] discusses it, but deals with it only when the optimum value of X_n is <u>strictly</u> <u>inside</u> the box $G = \{x: |x_i| \le B_i\}$. The methods of [1] can readily handle such problems whether or not the unconstrained optimum is in G. Assumptions are stated in Section 2. These are of the type used in [1] and are quite unrestrictive. In Section 3, it is shown that $\{X_n\}$ converges w.p.1 (under assumptions in Section 2) to the point in G which is <u>closest</u> to the optimum value. Incidentally, if the optimum is strictly interior to G, then the rate of convergence results in [5] hold. Section 4 deals with a formulation where $a_n \equiv \beta > 0$, a constant, and discusses some limit results of a 'weak convergence' nature, also using techniques from [1].

In many of the proofs in [1], it is assumed that the iterate sequence $\{X_n\}$ is bounded in some sense. Owing to the possible use of the projection algorithm (as in this paper), this boundedness assumption is hardly a restriction. This is one of the secondary points of this paper.

2. Assumptions for w.p.1. Convergence.

Define $m(t) = \max\{n; t_n \le t\}$, $t \ge 0$, where $t_n = \sum_{i=0}^{n-1} a_i$,

$= 0$, $t < 0$.

A1. There is a positive definite symmetric matrix R such that for each $\varepsilon > 0$ and some $T < \infty$

$$\lim_{n\to\infty} P\{\sup_{j\ge n} \max_{t\le T} \left| \sum_{i=m(jT)}^{m(jT+t)-1} a_i(\psi_i\psi_i' - R)\right| \ge \varepsilon\} = 0 \tag{2.1}$$

A2. There is a vector S such that (2.1) holds with $(\psi_n d_n - S)$ replacing $(\psi_i\psi_i' - R)$.

A3. (2.1) holds with $\psi_n\rho_n$ replacing $(\psi_i\psi_i' - R)$.

Note that (A1) - (A3) imply that

$$a_n[|\psi_i\psi_i'| + |\psi_n d_n| + |\psi_n\rho_n|] \to 0 \text{ w.p.1 as } n \to \infty.$$

Also, $\left|\sum_{i=m(t)}^{m(t+s)} a_i(\psi_i\psi_i' - R)\right| \to 0$ as $t \to \infty$, uniformly on bounded s-intervals w.p.1, and similarly for the cases of (A2), (A3). (See [1, Lemma 2.2.1]for the proof.)

Assumptions (A1) - (A3) are quite unrestrictive, [1, p.30-38] gives several ways of readily verifying them, and they hold in practical cases where (at least) the sequences $\{u_n, \rho_n, d_n\}$ are stationary. The reader is referred to the cited reference for more detail. We note only that the criteria for (A1) - (A3) in [1] can be weakened even further by a finer use of laws of large numbers and estimates of the type [1,(2.2.8)] which give bounds on $E \max_{n \le i \le N} \left|\sum_n^i \xi_j\right|^2$ in terms of bounds on the correlation function of $\{\xi_n\}$. (A1) - (A3) hold when $\{\psi_i\psi_i' - R\}$, etc., obey certain laws of large numbers or when their covariances go to zero sufficiently fast. E.g., if the processes are stationary and the covariances are summable, and $\sum(a_i \log_2 i)^2 < \infty$, then they hold. [1, Theorem 2.2.2].

3. Convergence w.p.1.

For any x, let $\pi_G(x)$ denote the nearest point on G to x. Then the truncated or projected form of (1.1) is

$$\tilde{X}_{n+1} = X_n - a_n\psi_n(\psi_n' X_n - d_n - \rho_n) \equiv X_n - h_1(X_n, \xi_n),$$

(3.1)

$$X_{n+1} = \pi_G(\tilde{X}_{n+1}),$$

where $\xi_n = (\psi_n, d_n, \rho_n)$.

Define $h(x) = -Rx + S$. and $\theta = R^{-1}S$. Then $h(x) = -R(x-\theta)$.

Define the projection $\bar{\pi}(h(x))$ of the vector field given by $h(x)$ onto G by

$$\bar{\pi}(h(x)) = \lim_{0<\Delta\to 0} [\pi_G(x+\Delta h(x))-x]/\Delta.$$

Then $\dot{x} = \bar{\pi}(h(x))$ is the 'constrained' or 'projected' flow corresponding to $\dot{x} = h(x)$. Basically, the limits of $\{X_n\}$ are those of the solution to $\dot{x} = \bar{\pi}(h(x))$, which in this case is the nearest point on G to θ.

Theorem 1. Assume (A1)-(A3). Then $\{X_n\}$ given by (3.1), converges w.p.1 to the closest point $\pi_G(\theta)$ in G to θ, the optimal value.

Proof. The various parts of the proof appear in [1, Theorem 5.3.1] (projection algorithm) and [1, Theorem 2.4.1 and 2.4.2] (a general unconstrained Robbins-Monro algorithm), and in order to avoid duplication we will merely put the pieces together.

By (A1) - (A3), $|X_{n+1}-X_n| \to 0$ w.p.1 and there is a sequence of positive real numbers $\{\gamma_n\}$, $\gamma_n \to 0$ as $n \to \infty$, such that $|X_{n+1}-X_n| \leq \gamma_n/2$ except for a finite number of terms, w.p.1. Let $I_n \equiv$ indicator of the set $\{|X_{n+1}-X_n| \leq \gamma_n/2\}$. Define

$$v_n^\gamma \equiv X_n - a_n h_1(X_n,\xi_n) I_n, \quad v_n \equiv X_n - a_n h_1(X_n,\xi_n) = \tilde{X}_{n+1},$$

$$\tau_n \equiv [\pi_G(v_n^\gamma) - v_n^\gamma], \qquad \phi_n \equiv (v_n^\gamma - v_n) + [\pi_G(v_n) - X_n](1-I_n).$$

Then

$$X_{n+1} = X_n - a_n h_1(X_n,\xi_n) + \tau_n + \phi_n. \tag{3.2}$$

Following [1, Theorem 5.3.1], define $\tau^0(\cdot)$ and $\phi^0(\cdot)$, resp., to be the <u>piecewise linear</u> functions on $[0,\infty)$ with values $\sum_{i=0}^{n-1} a_i\tau_i$ and $\sum_{i=0}^{n-1} a_i\phi_i$, resp., at time t_n. Let $\tau^n(\cdot) = \tau^0(t_n+\cdot) - \tau^0(t_n)$, $\phi^n(\cdot) = \phi^0(t_n+\cdot) - \phi^0(t_n)$. Let $\bar{X}^0(\cdot)$ and $\bar{\xi}^0(\cdot)$ denote the <u>piecewise constant</u> functions on $[0,\infty)$ which are equal to X_n and ξ_n, resp., on $[t_n,t_{n+1})$, where $t_n = \sum_{i=0}^{n-1} a_i$. Define $H_1^n(t) = \int_0^t h_1(\bar{X}^0(t_n+s), \bar{\xi}^0(t_n+s))ds$, and let $X^0(t)$ denote the <u>piecewise linear</u> function with value X_n at t_n and set $X^n(\cdot) = X^n(t_n+\cdot)$. Then $(X^n(0) = X_n)$

$$X^n(t) = X^n(0) + H_1^n(t) + \phi^n(t) + \tau^n(t).$$

<u>Fix</u> ω <u>not in one of the exceptional sets of zero probability of</u> (A1) - (A3). Then as in [1, Theorem 5.3.1], $\phi^n(\cdot) \to 0$ uniformly on bounded intervals as $n \to \infty$. Similarly (following

the same proof) if $\{H_1^n(\cdot)\}$ were equicontinuous, then so would $\{\tau^n(\cdot)\}$ and $\{X^n(\cdot)\}$ be.

Suppose for the moment that $\{H_1^n(\cdot)\}$ is equicontinuous. Then we can select a subsequence (also indexed by n) such that all $X^n(\cdot)$, $X^n(0)$, $H_1^n(\cdot)$, $\phi^n(\cdot)$, $\tau^n(\cdot)$ converge uniformly on bounded intervals, with limits denoted by $X(\cdot)$, $X(0)$, $H_1(\cdot)$, 0, $\tau(\cdot)$. Then

$$X(t) = X(0) + \int_0^t H_1(s)ds + \tau(t). \tag{3.3}$$

We need only prove the equicontinuity of $\{H_1^n(\cdot)\}$ and characterize $H_1(\cdot)$. This is done in the next two paragraphs. It will turn out that $H_1(s) = -RX(s) + S$. But, then the proof of [1, Theorem 5.3.1] implies that

$$\dot{X} = \bar{\pi}(-RX + S) = \bar{\pi}(-R(X-\theta)).$$

Define $f(x) \equiv (x-\theta)'(x-\theta)$. Then $\dot{f}(x) = (x-\theta)'\bar{\pi}(-R(x-\theta))$, and [1, Theorem 5.3.1] implies our theorem because the only points where $\dot{f}(x) = 0$ are $x = \theta$ and $x = \pi_G(\theta)$ (which equals θ when $\theta \in G$). The uniqueness of the solution to the limiting differential equation (for each $X(0)$), and the uniqueness of its limit point, imply that the particular fixed ω and the chosen subsequence are irrelevant.

Equicontinuity of $\{H_1^n(\cdot)\}$ for the fixed ω. The equicontinuity and the characterization of $H_1(\cdot)$ follow by use of the method of [1, Theorems 2.4.1 and 2.4.2]. If $\{\xi_n\}$ is a bounded sequence, then $\{H_1^n(\cdot)\}$ is obviously equicontinuous since $h_1(X^0(u),\xi^0(u))$ is uniformly bounded. Otherwise, use the method of [1, Theorem 2.4.2], whose conditions are implied by (A1) - (A3). We can write

$$X_{n+1} = X_n - a_n(\psi_n\psi_n' - R)X_n + a_n(\psi_n d_n + \rho_n\psi_n - S)$$

$$- a_n(RX_n - S) + \phi_n + \tau_n.$$

Uniform continuity on $[0,\infty)$ of the piecewise linear function with value $\sum_{i=0}^{n-1} a_i(\psi_i d_i + \rho_i\psi_i - S)$ at t_n follows from (A2,3) for our fixed ω. (See the second sentence below (A3).)

We only need show uniform continuity on $[0,\infty)$ of the piecewise linear function with value $\sum_{i=0}^{n-1} a_i|\psi_i\psi_i' - R|$ at t_n. Once this is done, $H_1^n(\cdot)$, the shifted sequence, will obviously be equicontinuous (the X_n coefficient, being bounded, is not important). Let $\psi_n = (\psi_{n0},\ldots,\psi_{n,r-1})'$. It is sufficient to show the uniform continuity for the linear interpolation† of $\left\{\sum_{j=0}^{n-1} a_j\psi_{ji}^2\right\}$ for each i. But this follows from (A1), which implies that $\sum_{j=0}^{n-1} a_j\psi_{ji}^2$ increases asymptotically as $\sum_{j=0}^{n-1} a_j R_{ii} = t_n R_{ii}$, since (see note below (A3))

$$\left|\sum_{i=m(t)}^{m(t+s)} a_n(\psi_{ni}^2 - R_{ii})\right| \to 0 \text{ uniformly on bounded } s\text{-intervals as}$$

$t \to \infty$.

---

† i.e., for the piecewise linear function with value $\sum_{j=0}^{n-1} a_j\,\psi_{ji}^2$ at t_n.

Characterization of the limit $H_1(\cdot)$. Let n still index the convergent subsequence. For our fixed ω, we need only show that

$$(3.4) \qquad \left|\sum_{j=m(t_n)}^{m(t_n+t)} a_j(\psi_j\psi_j'-R)X_j\right| \to 0 \quad \text{uniformly on bounded}$$

t-intervals as $n \to \infty$,

since it follows from this and (A2,3) that $H^n(\cdot) \to -RX(\cdot) + S$, as desired.

(3.4) can readily be done by following almost word for word the method of proof of a similar result in [1, Theorem 2.4.1], with the identifications $\theta(|X-X'|) = \text{const.}|X-X'|$ g_1 = constant and $g_2(\xi_j) = \sum_{i=0}^{r-1} \psi_{ji}^2$. Then (A1) implies (3.4). The argument in the reference shows that the $\{X_n\}$ fluctuations are 'slow enough' in comparison with those of the $(\psi_j\psi_j' - R)$ for the limit of (3.4) to be the same as it would be if X_n were a constant.

Q.E.D.

4. Constant $a_n \equiv \beta > 0$.

Theorems 4.3.1 and 6.2.3 present the analogs of Theorems 2.4.1 and 5.3.1, resp., under weaker conditions than (A1) - (A3) but where the convergence is in a weak sense. Instead of adapting them to the current projection problem, we merely formulate their use in another related and very important problem: algorithm (3.1) where $a_n \equiv \beta > 0$. Indeed, many, if not most, uses of (3.1) use constant coefficients $a_n \equiv \beta$, (at least for the 'tails' - once the 'transient' period is over). For small β, it might take quite a while for (transient period) X_n to get near its limit (θ or the projection of θ on G). We are concerned (more or less) with what occurs after this transient period. Let X_n^β denote the solution to (3.1) and define $X^\beta(\cdot)$ = piecewise linear interpolation (intervals β) of $\{X_n^\beta\}$. Let N_β be a sequence of integers (roughly defining the transient period, perhaps), such that $N_\beta \to \infty$ and $N_\beta\beta \to \infty$ as $\beta \to 0$.

A4. Let ξ_n represent $(\psi_n\psi_n' - R)$ or $\psi_n d_n - S$ or $\psi_n\rho_n$. Here $m(t) = [t/\beta]$ the integral part of t/β. Assume

$$\overline{\sup_{t\to\infty}}\, P\{\max_{s\leq T} \Big|\sum_{i=m(t)}^{m(t+s)} a\xi_i\Big| \geq \varepsilon > 0\} = 0, \quad \text{any } T < \infty,$$

each $\varepsilon > 0$.

This condition is discussed after the Theorem.

Theorem 2. Assume that $E|\psi_n\psi_n'|^2$, $E|\psi_n\rho_n|^2$ and $E|\psi_n d_n|^2$ are uniformly bounded, and assume (A4). Then $\{X^\beta(N_\beta\beta+\cdot)\}$ converges weakly (in the function space $C^r[0,\infty)$) to the constant function $\pi_G(\theta)$ as $\beta \to 0$, where $\pi_G(\theta)$ = nearest point to θ on G. In particular $X^\beta_{N_\beta\beta+k} \to \pi_G(\theta)$ in probability as $\beta \to 0$, each $k > 0$, and more strongly

$$\lim_{\beta\to 0} P\{\sup_{k\leq t/\beta} |X^\beta_{N_\beta\beta+k} - \pi_G(\theta)| \geq \varepsilon > 0\} = 0 \quad \text{for each } t$$

and $\varepsilon > 0$.

Thus for small β and large n, $\{X_n\}$ 'hovers' around $\pi_G(0)$ as desired.

The proof will not be given since it follows the general lines of the appropriate parts of [1, Theorems 4.3.1 and 6.2.3], which, in turn, are just weak convergence analogs of the theorems upon which Theorem 1 is based. To adapt the proofs of [1] to the present case, merely replace the shifted sequences $X^n(\cdot)$, etc. of [1] by $X^\beta(N_\beta\beta+\cdot)$, etc. If $E|\xi_i|^2$ is bounded as assumed above then, by [1, Theorem 4.1.1] (A4) holds if there are $R(i)$ such that $|E\xi_j\xi_{j+i}'| \leq R(i) \to 0$ as $i \to \infty$, a very weak condition indeed.

<u>Conclusions.</u> The projected iterate sequence converge w.p.1 to the closest point on G to the optimum (converges in probability in the weak convergence case).

REFERENCES

[1] H.J. Kushner, D.S. Clark, Stochastic Approximation Methods for Constrained and Unconstrained Systems, Springer, Berlin, 1978, Applied Math. Sciences Series #26.

[2] L. Ljung, 'Analysis of recursive stochastic algorithms', IEEE, Trans. on Automatic Control, AC-22, 1977, 551-575.

[3] A. Gersho, 'Adaptive equalization of highly dispersive channels for data transmission', B.S.T.J., 48, 1969, p.55-70.

[4] J. Proakis, 'Advances in equalization for intersymbol interference', Adv. in Comm. Syst., Vol 4, 1975, Academic Press, A.V.Balakrishnan, editor.

[5] H.J. Kushner, Hai Huang, 'Rates of convergence for stochastic approximation type algorithms', to appear SIAM J. on Control and Optimiz., Sept., 1979.